



Comparing national and global data collection systems for reporting, outbreaks of H5N1 HPAI

Matthew L. Farnsworth^{a,*}, Christopher Hamilton-West^b, Stephanie Fitchett^a,
Scott H. Newman^b, Stephane de La Rocque^{b,c}, Lorenzo De Simone^b,
Juan Lubroth^b, Julio Pinto^b

^a USDA-APHIS-VS-Centers for Epidemiology and Animal Health, 2150 Centre Avenue, Bldg. B, Mail Stop 2W4, Fort Collins, CO 80526, USA

^b Emergency Prevention System for Priority Animal and Plant Pests and Diseases/Global Early Warning and Response System (EMPRES/GLEWS), Animal Health Service, Animal Production and Health Division, Food and Agriculture Organization of the United Nations, Viale delle Terme di Caracalla, Rome 00153, Italy

^c Centre International de Recherche Agronomique pour le Développement (CIRAD), Campus International de Baillarguet, 34398 Montpellier Cedex, France

ARTICLE INFO

Article history:

Received 9 October 2009

Received in revised form 19 March 2010

Accepted 25 March 2010

Keywords:

H5N1 HPAI

Global database

Disease surveillance

Statistical analysis

Spatial scale

ABSTRACT

Determining if outbreak data collected by regional or international organizations can reflect patterns observed in more detailed data collected by national veterinary services is a necessary first step if global databases are to be used for making inference about determinants of disease maintenance and spread and for emergency planning and response. We compared two data sources that capture spatial and temporal information about H5N1 highly pathogenic avian influenza outbreaks reported since 2004 in four countries: Bangladesh, Egypt, Turkey, and Vietnam. One data source consisted of reports collected as part of each country's national veterinary services surveillance program, while the other data source included reports collected using the Emergency Prevention System for Priority Animal and Plant Pests and Diseases (EMPRES-i) global animal health information system. We computed Spearman rank-order correlation statistics to compare spatial and temporal outbreak distributions, and applied a space–time permutation test to check for consistency between the two data sources. Although EMPRES-i typically captured fewer outbreaks than detailed national reporting data, the overall similarity in space and time, particularly after 2006, reflect the ability of the EMPRES-i system to portray disease patterns comparable to those observed in national data sets. Specifically, we show that the two datasets exhibit higher positive correlations in outbreak timing and reported locations after 2006 when compared to December 2003 through 2006. Strengthening the capacity of global systems to acquire data from national and regional databases will improve global analysis efforts and increase the ability of such systems to rapidly alert countries and the international community of potential disease threats.

Published by Elsevier B.V.

1. Introduction

H5N1 Highly Pathogenic Avian Influenza (HPAI) is an example of a zoonotic disease with large scale impacts on

animal production, livelihoods, economies, wildlife, and public health (Rushton et al., 2005). The local severity of this disease has resulted in the establishment of national surveillance systems in several countries, while interest in larger scale outbreak patterns (e.g., regional or continental) has driven development of global information systems to collect, store and analyze outbreak information. One global system is the Emergency Prevention System for Priority

* Corresponding author. Tel.: +1 970 494 7293.

E-mail address: Matt.Farnsworth@aphis.usda.gov (M.L. Farnsworth).

Animal and Plant Pests and Diseases (EMPRES-i) global animal health information system (Martin et al., 2007b) developed by the Food and Agriculture Organization (FAO) of the United Nations. EMPRES-i has been tracking H5N1 HPAI since December 2003.

Global animal health information systems such as EMPRES-i serve many purposes. For instance, they improve access to outbreak data for disease analysis and support notification activities by monitoring and summarizing the global status of priority animal diseases and zoonoses, including H5N1 HPAI. Effective notification of disease outbreaks and the capacity for understanding potential spread to new areas is an essential pre-requisite for successful containment and control of epidemic diseases (Ben Jebara, 2004; Martin et al., 2007b; Pittman et al., 2007), including emerging zoonoses such as H5N1 Highly Pathogenic Avian Influenza (HPAI) and other transboundary animal diseases. From a public health perspective notification of animal outbreaks with a known zoonotic potential enables health authorities to warn at-risk populations of behaviors that should be avoided and implement control measures that can prevent human morbidity and mortality (Reed et al., 2003; Webster and Hulse, 2004). From an animal production point of view endemic diseases and epidemics in livestock negatively impact food security and the livelihood of the poorest, along with affecting international trade (Thiermann, 2004; Zepeda et al., 2005; Doyle and Erickson, 2006; Pinto et al., 2008). In some cases spill over of disease between livestock and wildlife populations might also be prevented if an outbreak notification system is in place and proper actions taken. Another important purpose served by global databases is that of providing a central source of data that may be analyzed to improve understanding of disease patterns and ecological mechanisms underlying disease spread and maintenance, and to identify how these mechanisms vary across countries and regions (Cecchi et al., 2008).

Information on several priority animal diseases (e.g., foot and mouth disease, Rift Valley fever, African swine fever, and rinderpest) is gathered through EMPRES-i (Martin et al., 2007a), and among them H5N1 HPAI was selected for this study for a variety of reasons. First, the dataset is relatively large, with 4860 outbreak reports between December 2003 and April 2009 covering the areas of interest; second, there were national data sets for H5N1 HPAI available from several countries for comparison to EMPRES-i; and third, H5N1 HPAI is a high priority animal disease globally, as indicated by the wide spread impacts noted above.

In addition to global databases several countries have implemented surveillance programs resulting in national databases, often containing more detailed information than global systems, maintained by each country's national veterinary service. National databases are built and populated using different technologies than EMPRES-i, for example, the use of digital pen systems and Short Message Service (SMS) Gateway technologies are increasing the capacity of national surveillance programs to collect disease data from the field and transmit it to a national database in real time. Additionally, in some countries where national surveillance systems have been developed

for H5N1 HPAI outbreak data are recorded at a finer spatial resolution than in EMPRES-i. Further, national databases tend to be more complete than EMPRES-i in terms of the number of outbreaks reported. In particular, in countries where H5N1 HPAI is endemic, such as Indonesia and Vietnam, national veterinary services are not required to report details of every outbreak to the OIE under the immediate notifications and follow-up reports in response to an exceptional disease event agreement, but instead maintain their own internal outbreak databases. Thus national databases may provide information about outbreaks that is unavailable in a global database. National databases maintained by each country's veterinary services were used in this study for its comparison to EMPRES-i data. When making our comparisons we assumed that national databases were more comprehensive in terms of the information and details regarding the number of outbreaks. Indeed, there were always more records in the national databases than in EMPRES-i for the time periods we compared.

For a system to be effective in improving the understanding of disease spread patterns and determinants the data on which the system relies must reflect those patterns in space and time. With this criterion in mind, the primary objective of this study was to determine the degree of concordance in spatial and temporal outbreak patterns of H5N1 HPAI between national veterinary services outbreak data, hereafter called "national" data, and EMPRES-i data and to determine both the extent of agreement and the degree to which that agreement has changed over time.

Specifically, we compare H5N1 HPAI outbreak data from national and EMPRES-i databases for four countries: Bangladesh, Egypt, Turkey, and Vietnam. These countries were chosen because they had a relatively large number of H5N1 HPAI outbreaks reported in both national and EMPRES-i data sources. Outbreak reports from these four countries constitute approximately 26% of all EMPRES-i H5N1 HPAI outbreaks reported between December 2003 and April 2009. Additionally, national data sets from these countries were available for at least some portion of this same time period. Indonesia and Thailand, which together represent an additional 44% of the EMPRES-i H5N1 HPAI data set, were also considered for this study, but were not used because there were political obstacles to using data from Indonesia, and national outbreak data from Thailand had already been directly integrated into EMPRES-i.

This is the first study we are aware of comparing agreement between national and global databases for an animal disease. Results of this analysis may help determine the extent to which EMPRES-i data might be used to develop an understanding of spatial and temporal dynamics of H5N1 HPAI at country, regional, and global scales.

2. Materials and methods

2.1. Data collection systems

EMPRES-i was developed to enable tracking and analysis of priority animal disease outbreaks across the globe. The system collects and stores information from multiple sources which can be classified as formal and informal. Formal data sources include reports from the

World Organization for Animal Health (OIE), the European Commission, the World Health Organization, national authorities, FAO country or regional projects, FAO field missions, FAO field officers, partner Non-Governmental Organizations, laboratories and reference centers, and other FAO collaborators. Informal sources include media reports from within countries or reports disseminated through, for instance, the Global Public Health Intelligence Network and ProMED. All information is followed by disease-tracking officers within FAO networks until the disease event is confirmed or denied. In addition, EMPRES-i staff verifies, and when necessary correct, the location and timing of outbreaks reported by each country to the formal data sources noted above using a variety of mapping tools (e.g., Google Earth) and by contacting national veterinary services staff. The same tools and official contacts are used to verify and correct data obtained from informal sources. Collecting and verifying data from these various sources allows EMPRES-i to deliver comprehensive situation updates on outbreaks involving priority animal diseases (Martin et al., 2007a) such as H5N1 HPAI, which may be unavailable in raw form from international reporting entities. The primary sources for EMPRES-i data specific to outbreaks of H5N1 HPAI are official sources that include national governments, official web sites, FAO projects and OIE reports.

For each outbreak, the EMPRES-i data system includes compulsory fields such as observation date, reporting date, and a geo-referenced location. In the case of H5N1 HPAI, the observation date represents the closest date to the recognition of clinical signs for an outbreak, and the reporting date reflects the confirmation and report of the outbreak event to official international organizations, for example OIE. The date on which an outbreak report is added to the EMPRES-i database, known as the insertion date, is automatically recorded as well. When comparing data sets, we used the observation date from EMPRES-i since this corresponds to the best estimate of the initiation date of an outbreak event. Depending on the country, the location of an outbreak may have been recorded exactly (using GPS technology), but was more often approximated by the centroid of the Administrative level 1, 2, or 3 unit in which the outbreak was believed to have occurred. Administrative units at lower levels are composed of several higher level units (e.g., one Administrative 1 unit is composed of multiple Administrative 2 units). For example, in Vietnam Administrative 1, 2, and 3 levels correspond to provinces, counties, and communes, respectively. Administrative boundary data used throughout the analysis were acquired from FAO's Global Administrative Unit Level GIS (FAO, 2006) data files, with coordinates specified by the World Geodetic System (WGS84).

Updates in the development and structure of the EMPRES-i database after 2006 included the implementation of standard operating procedures for disease tracking activities, which resulted in greater effort being placed on obtaining, verifying, and validating outbreak reports. Other changes included an increase in the breadth of species represented in the database, such as wildlife, and quality control on data entry and validation by automating checks between the reported location of outbreak coordinates

(e.g., latitude and longitude) and the reported administrative level and specific unit. Also a set of mandatory fields were included to provide a minimum standard for the information that must be included for each reported outbreak. As progressive development of the system is expected to correspond with improvements in the quality of the data, where possible, data analyses were conducted for two time periods that closely approximate the dates of major changes to the EMPRES-i system: before January 1, 2007, and since January 1, 2007. The goal of this comparison was to examine changes in the ability of EMPRES-i to reflect national data from each of the selected countries.

National level H5N1 HPAI outbreak data included reports collected directly from field veterinarians conducting surveillance, and in general contained more detailed information on disease outbreaks than what was reflected in the mandatory fields provided by EMPRES-i. National datasets were acquired from the national veterinary services of Bangladesh, Egypt, Turkey, and Vietnam; all of which have had surveillance activities enhanced by FAO technical projects that support specific programs aimed at strengthening disease surveillance. For example, in Egypt FAO facilitated the use of Participatory Disease Surveillance approaches, while in Bangladesh FAO assisted in implementing the use of SMS gateway technology to report mortality and morbidity in poultry by animal health workers in the field. All records included a field reflecting the date when an outbreak was observed. Outbreak locations were recorded as the centroid of the lowest administrative level possible. For Bangladesh national data, the centroids of sub-districts (i.e., Administrative level 3) were provided, while for Egypt, Turkey, and Vietnam, the geographic center of outbreak villages were provided.

Since national data sets and EMPRES-i were not necessarily recorded at the same spatial resolution we standardized both data sets by aggregating each one to the Administrative 1 and 2 levels prior to making comparisons. Thus, our inference related only to these two levels of aggregation. Temporally, each national data set spanned a different time period, and all are different from the December 2003 to April 2009 period covered by the EMPRES-i data used here. National datasets were provided to FAO "as is" and so we could only make a valid comparison by using EMPRES-i data from the same date range covered by the national databases. Since we did not have information on when active national surveillance started or ended in each country, the range of dates over which national data was compared to EMPRES-i data was determined by taking the date of the first and last reported outbreak in each national database as the initial and final date for comparison with EMPRES-i. For Egypt, Turkey and Vietnam, the date range available in the national data was a subset of what was available in EMPRES-i, and for Bangladesh, the date range began in 2007 (much later than EMPRES-i) and extended slightly longer than what was available from EMPRES-i.

2.2. Statistical analysis

Descriptive statistics were computed for the difference between observation date and reporting date for all H5N1 HPAI data in EMPRES-i.

Time series plots of the number of outbreaks per week in EMPRES-i and national data sets were overlaid for each of the four countries to visually compare the outbreaks recorded in the two data sources. As a measure of the similarity in these pairs of time series Spearman rank-order correlations between weekly time series from the two data sources were calculated for all four countries using the range of outbreak observation dates common to both data sets. That is, for each week of the time window common to national data and EMPRES-i, the number of reports from the national database each week (x_s) were paired with the number of reports each week from EMPRES-i (y), and the rank-order correlation was computed for the paired data. The overlap in outbreak observations were: 1 March 07–19 April 09 for Bangladesh, 17 February 06–4 January

the family of Monte Carlo permutation procedures (Efron and Tibshirani, 1993) developed to compare a statistic computed from observed data to a distribution of statistics based on randomizing the observed data many times. A p -value was calculated by determining where the observed test statistic fell within the distribution of randomization test statistics.

Specifically, if location and week of outbreak are provided in the form $(\mathbf{s}_{Ei}, t_{Ei})$, for $i = 1, \dots, n_E$ from EMPRES-i and $(\mathbf{s}_{Ci}, t_{Ci})$, for $i = 1, \dots, n_C$ from the national data, and $\mathbf{d}(\mathbf{s}_{Ei}, \mathbf{s}_{Ej})$ (respectively, $\mathbf{d}(\mathbf{s}_{Ci}, \mathbf{s}_{Cj})$) gives the Euclidean distance between locations \mathbf{s}_{Ei} and \mathbf{s}_{Ej} (respectively, \mathbf{s}_{Ci} and \mathbf{s}_{Cj}), then the test statistic for the space–time SMRPP was the weighted average of the “mean pairwise distance in space and time” for the two data sources, given by:

$$\delta = \frac{\frac{2}{n_E-1} \sum_{i < j} \sqrt{\mathbf{d}(\mathbf{s}_{Ei}, \mathbf{s}_{Ej})^2 + (t_{Ei} - t_{Ej})^2} + \frac{2}{n_C-1} \sum_{i < j} \sqrt{\mathbf{d}(\mathbf{s}_{Ci}, \mathbf{s}_{Cj})^2 + (t_{Ci} - t_{Cj})^2}}{n_E + n_C} \quad (1)$$

08 for Egypt, 25 November 05–31 March 06 for Turkey, and 29 March 04–18 March 09 for Vietnam. Non-parametric Spearman rank-order correlation was computed because the distributions of weekly reports are non-normally distributed counts. Confidence intervals at the 95% level were generated for the Spearman correlations by bootstrapping the observed data 10,000 times.

Among the four countries considered, only Vietnam and Egypt had long enough time series to compare changes in the EMPRES-i system's ability to track national data before and after improvements were made to the EMPRES-i system beginning in January 2007. Data for Vietnam and Egypt were divided into two time periods, an early time period that included outbreaks through the end of 2006, and a late time period that included outbreaks since January 1, 2007.

EMPRES-i and national outbreak reports were labeled according to which Administrative 2 unit each outbreak was reported to have occurred in for the purposes of comparing the two data sets at that spatial resolution. In addition, outbreak data were further classified at the more coarse Administrative 1 level to examine how agreement between the two data sources changed as a function of increased spatial aggregation. For each country, maps were generated displaying the distribution of outbreaks for both data sets at the first and second administrative unit level to visually compare spatial patterns between the two data sources at two spatial scales. Spearman rank-order correlations between the numbers of reports in each administrative unit were then computed for all four countries at both spatial scales. Confidence intervals for the correlations were calculated by bootstrapping, as above. As with the time series data, the spatial correlations for Vietnam and Egypt were calculated separately for data reported before and since January 1, 2007 (early and late time period, respectively) to quantify the effect of changes in the development of EMPRES-i information system over time.

Finally, a space–time multi-response permutation procedure (SMRPP; Merton et al., submitted for publication) was applied to test for concordance between both data sources for each country in terms of spatial, temporal, and spatio-temporal distributions. The SMRPP test comes from

The statistic for the spatial SMRPP was the same as that given in Eq. (1), but with only the $\mathbf{d}(\mathbf{s}_{Ei}, \mathbf{s}_{Ej})^2$ and $\mathbf{d}(\mathbf{s}_{Ci}, \mathbf{s}_{Cj})^2$ terms included under the radicals, while the statistic for the temporal SMRPP included only the $(t_{Ei} - t_{Ej})^2$ and $(t_{Ci} - t_{Cj})^2$ terms under the radicals. Testing proceeded by first evaluating the statistic for the observed data. For each randomization, the data source indices (the E s and C s in the first of the subscripts) were randomly permuted among the $n_E + n_C$ observations, and the statistic was computed for the permuted location–time values. The randomization process and statistic computation were repeated many times, and the p -value for the test was calculated as the proportion of all computed statistics (the observed one and all the randomized ones) that were less than or equal to the observed statistic.

For each country, before applying the SMRPP tests, data were first aligned temporally by only including date ranges common to both EMPRES-i and national databases as above. For location, we used the locations provided in each data source (coordinates of villages or other administrative centroids) and for time, we used the observation week of the outbreak. The data for Vietnam and Egypt were again divided into two time periods, before and after January 1, 2007, to test the null hypothesis of no differences in the shapes of the space, time, or space–time distributions of EMPRES-i outbreaks relative to national data sets for each time period. For each comparison we ran 1000 Monte Carlo simulations of the data. While assessing the spatial and temporal correlations provided a relatively coarse measure of agreement between the data sources, the more sensitive SMRPP tests compared the shapes of the spatial, temporal and spatio-temporal distributions of the data sets. Greater differences in distributional shapes between data sets resulted in smaller p -values for this test.

3. Results

3.1. Descriptive statistics

The distribution of the number of days between the observation date and the reporting date in EMPRES-i

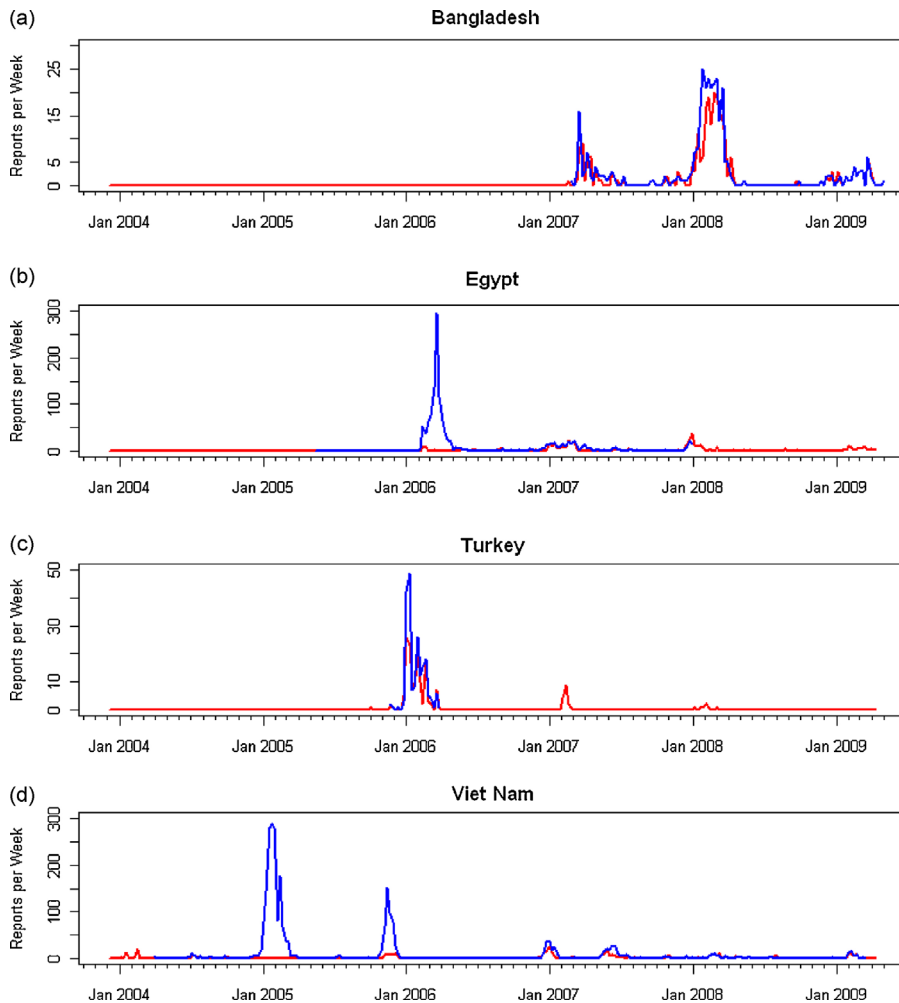


Fig. 1. (a, b, c, d) Time series plots comparing EMPRES-i outbreak reports (red) with national outbreak reports (blue). Note that large outbreaks in Egypt and Vietnam prior to 2007 were not captured by EMPRES-i. Improved matching of EMPRES-i with national data is seen after 2006.

for H5N1 HPAI reports of domestic poultry infection, regardless of country, was highly skewed. For outbreaks occurring in 2006 or earlier the 50th, 75th, and 90th percentiles were 0, 2, and 21 days, respectively. These same percentiles for the 2007 and later period were 0.5, 4, and 15 days. Finally, for outbreaks reported by EMPRES-i in 2009 the 50th, 75th, and 90th percentiles of the distribution were 2, 4, and 7 days, respectively, reflecting a substantial decrease in the variance within the upper tails of these distributions over time.

3.2. Time-series analysis

Comparison of time series (Fig. 1) between EMPRES-i and national reports for Vietnam and Egypt revealed discrepancies between the two databases. Specifically, there were large numbers of outbreaks reported in the national data for Egypt in 2006, but only very small numbers reported in EMPRES-i for the same time period. However,

there was greater similarity in the time series for outbreaks in 2008 and 2009. A similar pattern was seen in the Vietnam comparison, with large numbers of outbreaks in the winters of 2005 and 2006 reported in national data sets, with a much smaller number reported in the EMPRES-i database. As with the Egypt data, the two Vietnam data sources exhibited greater similarity in later time periods. Outbreaks reported in Turkey and Bangladesh showed good agreement between EMPRES-i and national data sets for the entire time period compared.

Correlations (Table 1) among time series captured the high degree of concordance between the EMPRES-i system and national data for Bangladesh ($r = 0.88$, C.I. = (0.83, 0.93)) and Turkey ($r = 0.93$, C.I. = (0.81, 0.98)). Also reflected were the positive changes made to EMPRES-i after 2006 with a doubling of the correlation in time series for Vietnam, from 0.43 (C.I. = (0.30, 0.56)) to 0.87 (C.I. = (0.81, 0.92)), and a greater than five-fold increase in the correlations for Egypt, from 0.14 (C.I. = (-0.10, 0.37)) to 0.74 (C.I. = (0.56, 0.86)).

Table 1

Spearman rank-order correlations and bootstrapped 95% confidence intervals (L_{CI} , U_{CI}) in weekly time series for early (through December 2006), and late outbreaks (since January 2007). Analysis includes reports recorded in the time window common to both national and EMPRES-i data sources for Vietnam (29 March 04–18 March 09), Egypt (17 February 06–4 January 08), Turkey (25 November 05–31 March 06), and Bangladesh (1 March 07–19 April 09). All reports with valid outbreak dates were used even if spatial information was not available. Note the improvement in time series correlations for Egypt and Vietnam from early to late time periods.

Time Period	Correlation (L_{CI} , U_{CI})	Country reports	EMPRES reports
Vietnam			
Early	0.43 (0.30, 0.56)	2098	120
Late	0.87 (0.81, 0.92)	395	274
Egypt			
Early	0.14 (−0.10, 0.37)	1003	51
Late	0.74 (0.58, 0.86)	283	236
Turkey			
Early	0.93 (0.81, 0.98)	201	137
Late	No data	0	0
Bangladesh			
Early	No data	0	0
Late	0.88 (0.83, 0.93)	323	259

3.3. Spatial pattern analysis

Example maps (Figs. 2 and 3) of Vietnam comparing the proportion of reported outbreaks reported in EMPRES-i to the proportion reported in national data for outbreaks reported before and after 2007, aggregated at the Administrative 1 and Administrative 2 levels, illustrated the higher correlation between EMPRES-i and national data in the spatial domain during the later time period. In both Vietnam and Egypt (Table 2), and at both levels of spatial aggregation, the correlation between the number of reports in EMPRES-i and in national data increased, particularly at the Administrative 1 level where the correlation for Vietnam went from 0.66 (C.I. = (0.51, 0.75)) in the early period to 0.90 (C.I. = (0.82, 0.95)) in the later period and the correlation for Egypt went from 0.58 (C.I. = (0.28, 0.81)) to 0.90 (C.I. = (0.78, 0.95)). For all four countries the numbers of outbreaks reported in EMPRES-i exhibited strong correlations with the numbers of outbreaks reported in national sources (Table 2) at the Administrative 1 level, ranging from 0.65 to 1.00, while the Administrative 2 level correlations were smaller and more variable, ranging from 0.05 to 0.82.

3.4. Space–time tests

Results from testing for spatial, temporal, and spatio-temporal consistency between EMPRES-i and national reports using the SMRPP procedure are presented in Table 3. For Turkey and Bangladesh, SMRPP distinguished no significant difference between the space, time, or space–time patterns in the EMPRES-i and national data, suggesting strong concordance between the two data sets over the periods in which the databases overlapped.

For Egypt and Vietnam, as would be expected from the time series plots (Fig. 1), the temporal patterns in EMPRES-i and national data were statistically different for the early time period. Likewise, the spatial and spatio-temporal patterns in the two data sources were significantly different for Egypt and Vietnam in the early time period. In the later time period, the SMRPP test resulted in no significant difference in the spatial distributions of reports, again suggesting increased concordance between national and EMPRES-i data in later periods. On the other hand, in contrast to the reasonably strong correlations seen in the time series for Vietnam and Egypt in the later time period, the

Table 2

Spearman rank-order correlations and bootstrapped 95% confidence intervals (L_{CI} , U_{CI}) between the number of national and EMPRES-i reports at the Administrative level 1 and 2 scales for early (December 2003 through December 2006) and late outbreaks (January 2007 through April 2009). Also reported for each country are the numbers of Administrative level 1 and 2 units. The analysis includes all reports for which valid spatial information was available.

Time period	Admin 1 correlation (L_{CI} , U_{CI})	Admin 2 correlation (L_{CI} , U_{CI})	Admin 1 units	Admin 2 units
Vietnam				
Early	0.66 (0.51, 0.75)	0.35 (0.27, 0.42)	61	497
Late	0.90 (0.82, 0.95)	0.50 (0.43, 0.58)		
Egypt				
Early	0.58 (0.28, 0.81)	0.05 (−0.06, 0.15)	26	289
Late	0.90 (0.78, 0.95)	0.31 (0.22, 0.41)		
Turkey				
Early	0.81 (0.71, 0.90)	0.73 (0.67, 0.79)	80	926
Late	No data	No data		
Bangladesh				
Early	No data	No data	6	64
Late	1.00 (NA, NA)	0.82 (0.71, 0.90)		

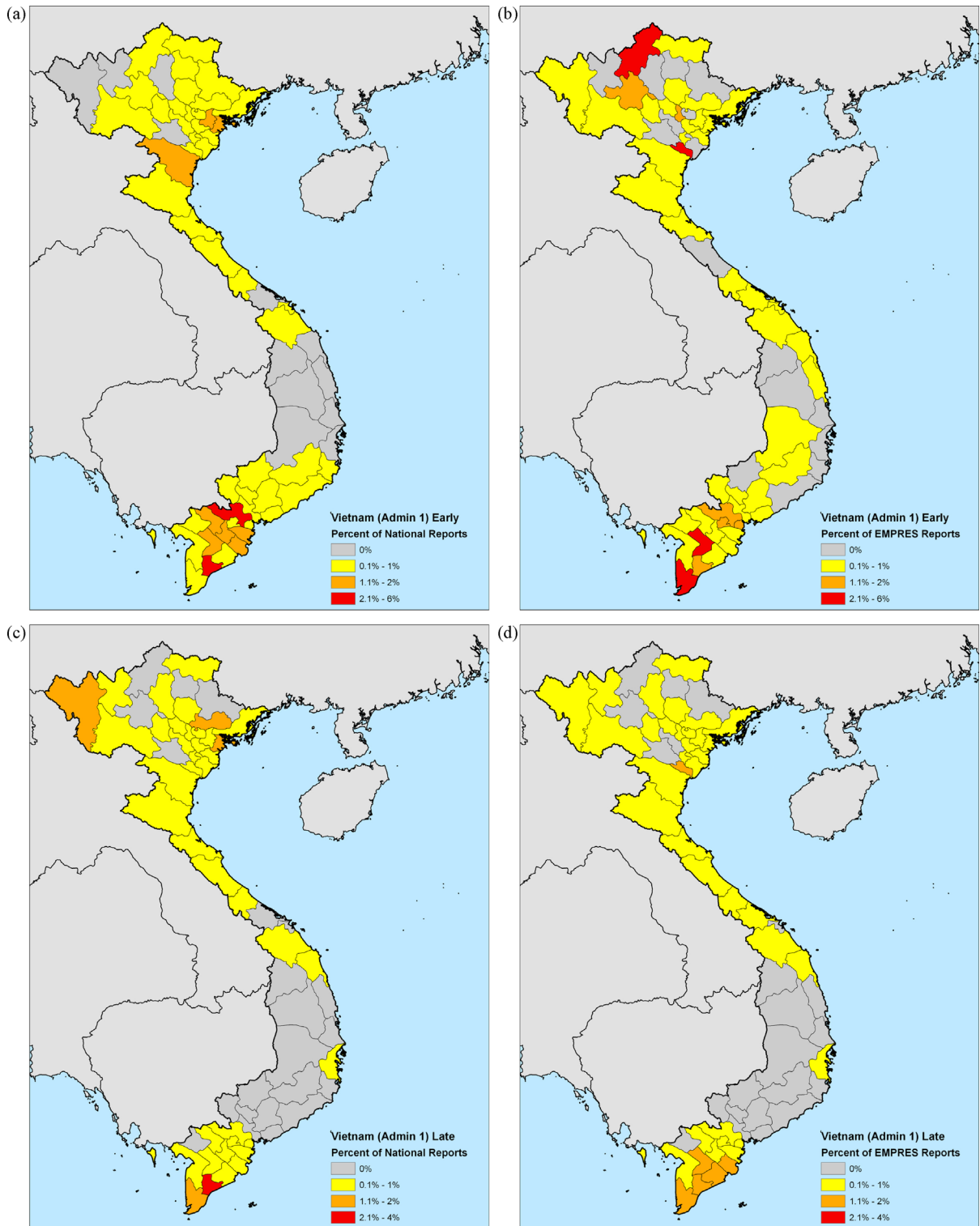


Fig. 2. (a, b, c, d) Comparison of EMPRES-i and national numbers of outbreaks reported at the Administrative-1 level for Vietnam during early (through December 2006) and late outbreaks (since January 2007): (a) national, early, (b) EMPRES-i, early, (c) national late, and (d) EMPRES-i, late. Map colors represent ranges for the percentage of all reports for a given time period which occurred in each Administrative-1 unit. The correlation in ranked outbreak numbers is 0.65 for the early period and 0.90 for the late period.

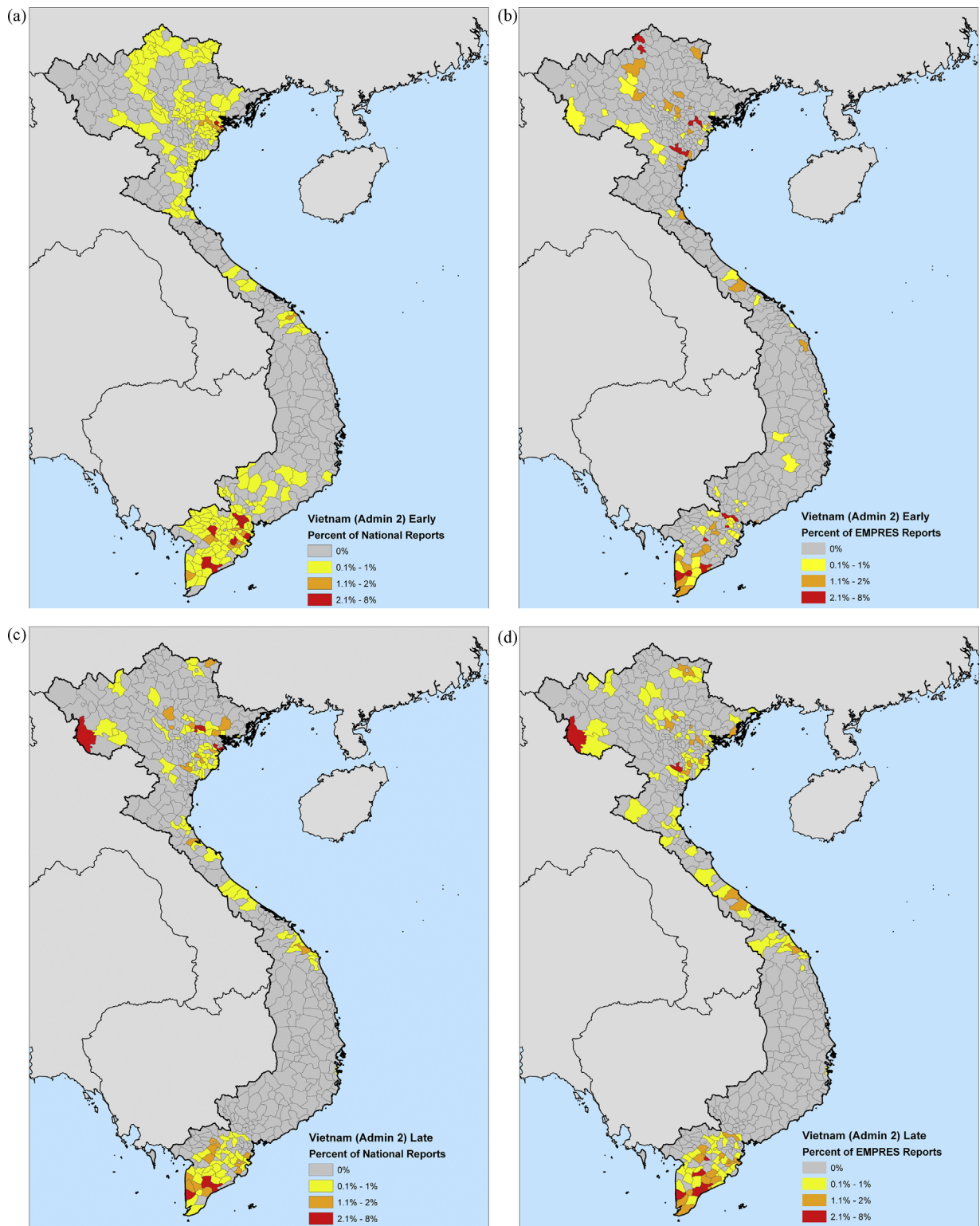


Fig. 3. (a, b, c, d) Comparison of EMPRES-i and national numbers of outbreaks reported at the Administrative-2 level for Vietnam during early (through December 2006) and late outbreaks (since January 2007): (a) national, early, (b) EMPRES-i, early, (c) national late, and (d) EMPRES-i, late. Map colors represent ranges for the percentage of all reports for a given time period which occurred in each Administrative-2 unit. The correlation in ranked outbreak numbers is 0.35 for the early period and 0.50 for the late period.

Table 3

Space–time Multi-Response Permutation Procedure (SMRPP) *p*-values for early (through December 2006) and late outbreaks (since January 2007), along with the total number of reports for each country recorded in each data source and time period. Large *p*-values (e.g., $p \geq 0.05$) indicate little difference between the distributions of EMPRES-i and national data for the dimension being considered. The analysis includes all reports with valid spatial information in the time window common to both national and EMPRES-i data sources.

Time period	<i>p</i> -Value space	<i>p</i> -Value time	<i>p</i> -Value space–time	Country reports	EMPRES reports
Vietnam					
Early	0.001	0.001	0.001	2076	120
Late	0.197	0.001	0.001	353	274
Egypt					
Early	0.001	0.001	0.001	996	51
Late	0.423	0.001	0.001	283	230
Turkey					
Early	0.219	0.663	0.270	201	137
Late	No data	No data	No data	0	0
Bangladesh					
Early	No data	No data	No data	0	0
Late	0.600	0.372	0.726	309	259

temporal SMRPP test indicated significant differences in the shapes of the weekly time series of outbreak reports between the data sources over this period. The SMRPP test also indicated significant differences in the shapes of the space–time distributions of reports for Vietnam and Egypt in the later time period.

4. Discussion

The goal of this study was to determine the level of agreement between outbreak data captured through the EMPRES-i disease tracking system and outbreak data contained in national data sets, particularly before and after improvements to EMPRES-i were implemented in January of 2007. Developing high quality data systems and determining if outbreak data collected by regional or international organizations can adequately reflect general patterns seen in data collected by national veterinary services are important first steps if global databases are to be used for making inference about disease spread patterns and processes (Martin et al., 2007b), especially for priority animal diseases like H5N1 HPAI. Disease data collected at a national level are typically considered to have greater reliability in terms of spatial and temporal reporting accuracy; thus the extent to which data collected by EMPRES-i reflects patterns in disease outbreak data collected by national surveillance programs is of great interest if EMPRES-i is to be used to quantify disease dynamics at country, regional, and global scales, and as an emergency preparedness and response system.

The EMPRES-i system has been improved over time, and we believe the improvements are reflected by an increased ability to track national outbreak data. Specifically, this study showed that correlations in time series and spatial correlations in reports aggregated at the Administrative 1 and 2 levels all improved over time for both Vietnam and Egypt, the two countries with a complete dataset of H5N1 HPAI outbreaks covering both the early and late period of interest. Aggregating data to the larger Administrative 1 level necessarily resulted in an increase in the correlations and reduction in variance

Prior to using these data for more advanced epidemiological analysis it is important to consider the spatial resolution of outbreak reports for a country, as well as potential variations in the spatial resolution of reporting among countries and the desired scale of epidemiological inference. Therefore, it will often be necessary to assign outbreak reports to a common spatial resolution (e.g., Administrative level 2) prior to analysis that includes more than one country. This work illustrates the effect of data aggregation on the correlation between EMPRES-i and national data sets and we recommend that future analyses using EMPRES-i data consider these results along with the considerations noted above.

Before 2006, there were a large number of outbreaks in Vietnam and one in Egypt (see Fig. 1) that were captured in national data sets, but not reflected in EMPRES-i due to the near absence of reporting by Vietnam and Egypt to OIE or other official international sources at that time. The fact that large numbers of cases were reported to national databases, but not captured by EMPRES-i, suggests that there may have been obstacles to getting some outbreaks reported to EMPRES-i. Such problems may have arisen, for instance, due to a lack of awareness about the need and importance of reporting to the international community. Other contributing factors might have been the absence of a formal mechanism for reporting local outbreaks contained within national databases to regional and international organizations, a failure of surveillance systems to gather disease outbreak data in standard formats (Salman et al., 2003; Thiermann, 2005), unwillingness to report outbreak events to the international community (Zepeda et al., 2005), or that animal diseases are not a priority for the national authorities and other stakeholders in some countries (Vallat et al., 2006). Some of these aspects remain obstacles to improving disease reporting from affected countries.

That said, using the national databases from Vietnam and Egypt allowed us to demonstrate substantial improvements in consistency between EMPRES-i and national data between the early and late period in terms of tracking disease patterns. Since January 2007, EMPRES-i and the national databases for Egypt and Vietnam have shown

strong agreement in spatial reporting at the relatively coarse Administrative 1 scale, and to a lesser degree at the Administrative 2 level. They have also shown consistency in temporal reporting as indicated by strong correlations between time series. Although SMRPP tests indicate no significant differences in the spatial distributions for these two countries after 2006, there were significant differences in the temporal and spatio-temporal distributions of reports, not only through 2006, but also in the later period beginning in 2007. SMRPP is quite sensitive to relatively small differences in the shapes of distributions being compared (Merton et al., submitted for publication), but such differences may not be critical to analyses focused on more general patterns and large scale dynamics. When SMRPP tests do not detect differences in distributions, as was the case for Turkey and Bangladesh in space, time, and space–time, EMPRES-i provides a highly reliable proxy for national data.

The increased concordance between national and EMPRES-i data sets over time can be attributed to both improvements in the logistics of data collection and stronger cooperation in reporting by national veterinary services. Standardizing the format for data collection and transfer included the development of a standard Excel template for reporting disease outbreak information to EMPRES-i. In Egypt and Vietnam, greater reporting cooperation was due, at least in part, to FAO supporting field officers who worked with veterinary services to improve surveillance and reporting. As national veterinary services become more aware of the importance of reporting outbreaks to the international community they tend to report more consistently. This is one reason why there were fewer discrepancies seen between EMPRES-i and national data sources during the second phase of development of EMPRES-i after January 2007. There still exist reporting differences between national and EMPRES-i data, including misidentification of the specific administrative unit where an outbreak occurred, and changes in epidemiological units, that is location data related to changes in administrative levels that need to be adjusted and harmonized between national databases and GAUL standards in FAO. Finally, outbreak events that warrant multiple reports in a national data set are sometimes grouped into a single report in EMPRES-i. Despite these differences, the data sources have become less independent over time, which is a desirable trend since ideally all detailed national data should be incorporated into EMPRES-i whenever possible. This situation should continue to improve because transparency of disease reporting for H5N1 HPAI has increased over time, particularly from endemic countries where the risk to humans is ever present. Unlike official reporting bodies EMPRES-i verifies and incorporates reports from a broad array of sources, not just data national veterinary services report to official international, regional, and global animal health information platforms. This capacity to combine information coming from official sources and projects in the field related to surveillance of H5N1 HPAI makes EMPRES-i a unique database for animal disease outbreak and surveillance data.

In contrast to the situation for Egypt and Vietnam where there were large discrepancies between EMPRES-

i and national data sources through 2006, for Bangladesh, which had outbreaks beginning in 2007, and Turkey, which experienced all outbreaks prior to 2007, the information collected in EMPRES-i appeared to reflect national data systems reporting for the entire time period in which outbreak data were available for comparing of both data systems. This strong agreement is due to several conditions that existed in these countries during the time periods examined. First, because of strong international support, which included training in disease outbreak investigations and developing laboratory capacity for disease detection and reporting systems, these countries were able to enhance their surveillance systems for H5N1 HPAI (Vallat et al., 2006). Second, standardized disease reporting procedures were established very early in both countries, including consistent reporting at defined administrative levels.

In terms of the timeliness of reporting, our results show an overall decrease in the time between outbreak and reporting date, along with a reduction in the variance, as the EMPRES-i system has evolved over time. This trend is encouraging and reflects improved data collection and reporting capabilities of EMPRES-i, along with suggesting an overall increase in the transparency and reporting capabilities of many countries affected by H5N1 HPAI. These results support increased use of EMPRES-i data for developing and prioritizing disease response activities, as well as its use as an early warning tool for identifying areas at risk for new incursions. This use of EMPRES-i may encourage local animal health authorities to increase biosecurity and surveillance activities in locations near an ongoing outbreak.

Considerable variation between countries in their veterinary services capacities, surveillance protocols, standards, and the level of reporting to global databases (Angot, 2009), makes analysis of data generated from these different countries challenging. It should be recognized that outbreaks of H5N1 HPAI, as is the case with any animal disease, are commonly underreported across all types of production systems, particularly in the commercial poultry sector for many countries. Additionally, active surveillance in countries where the disease is endemic is frequently low or absent, leaving surveillance systems with a limited capacity to detect and respond to an outbreak situation. Further, compensation for culled animals, when it exists, is often inadequate (Hadorn et al., 2008), further reducing the likelihood of outbreak reporting. Finally, reporting of transboundary diseases in wildlife is very limited, and in the case of H5N1 HPAI, under-reporting is common. Most often proper epidemiological outbreak investigations are not conducted or coordinated with wildlife resource ministries, leading to incomplete information being provided to international organizations on species affected. Despite the many obstacles, in recent years there has been general improvement in disease awareness resulting in greater levels of reporting by affected countries.

The identification of general patterns of disease spread in space and time, particularly for priority animal diseases such as H5N1 HPAI that persist and spread across large regions, thereby affecting several countries simultaneously, is essential for disease management (Fèvre et al., 2006). EMPRES-i is an important reference source for

disease outbreak information and is used by international agencies (e.g., Institute of Medicine and National Research Council) and national governments tracking the global status of disease outbreaks (Martin et al., 2007b). Since external users of EMPRES-i include research institutions and scientists that regularly request data for analysis of priority animal diseases, we have highlighted in this work both its utility and limitations for making inferences about disease patterns and processes. Future work should involve use of EMPRES-i data to explore variations in H5N1 HPAI patterns and dynamics at multiple spatial and temporal scales, including identifying variations in disease patterns and dynamics between countries and identifying potential mechanisms driving observed patterns of persistence and spread.

Conflict of interest statement

The authors claim no conflict of interest.

Acknowledgements

We thank Kristin King for her assistance in developing time-series plots and maps and the United States Department of Agriculture, International Services and Veterinary Services, and the FAO EMPRES program for supporting this work. We also wish to thank two anonymous reviewers for helpful comments and suggestions.

References

- Angot, J.-L., 2009. The governance of veterinary services and their role in the control of avian influenza. *Rev. Sci. Tech. Off. Int. Epiz.* 28, 397–400.
- Ben Jebara, K., 2004. Surveillance, detection and response: managing emerging diseases at national and international levels. *Rev. Sci. Tech. Off. Int. Epiz.* 23, 709–715.
- Cecchi, G., Ilemobade, A., Le Brune, Y., Hogerwerf, L., Slingenbergh, J., 2008. Agro-ecological features of the introduction and spread of the highly pathogenic avian influenza (HPAI) H5N1 in northern Nigeria. *Geospatial Health* 3, 7–16.
- Doyle, M.P., Erickson, M.C., 2006. Emerging microbiological food safety issues related to meat. *Meat Sci.* 74, 98–112.
- Efron, B., Tibshirani, R., 1993. *An Introduction to the Bootstrap*. Chapman & Hall, New York.
- FAO, 2006. *The Global Administrative Unit Layers (GAUL) Technical Aspects*. GAUL/Doc 01, Italy, p. 16.
- Fèvre, E.M., Bronsvoort, B.M.D.C., Hamilton, K.A., Cleaveland, S., 2006. Animal movements and the spread of infectious diseases. *Trends Microbiol.* 14, 125–131.
- Hadorn, D., Haracic, S., Stark, K., 2008. Comparative assessment of passive surveillance in disease-free and endemic situation: example of *Brucella melitensis* surveillance in Switzerland and in Bosnia and Herzegovina. *BMC Vet. Res.* 4, 52.
- Martin, V., De Simone, L., Lubroth, J., 2007a. Geographic information systems applied to the international surveillance and control of transboundary animal diseases, a focus on highly pathogenic avian influenza. *Vet. Ital.* 43, 437–450.
- Martin, V., von Dobschuetz, S., Lemenach, A., Rass, N., Schoustra, W., DeSimone, L., 2007b. Early warning, database, and information systems for avian influenza surveillance. *J. Wildl. Dis.* 43, S71–76.
- Merton, A.A., Hoeting, J.A., Webb, C.T., submitted for publication. Distribution-free comparison of multiple-response geo-referenced data. *J. Am. Stat. Assoc.*
- Pinto, J., Bonacic, C., Hamilton-West, C., Romero, J., Lubroth, J., 2008. Climate change and animal diseases in South America. *Rev. Sci. Technol.* 27, 599–613.
- Pittman, M., Laddomada, A., Freigofas, R., Piazza, V., Brouw, A., Brown, I.H., 2007. Surveillance, prevention, and disease management of avian influenza in the European Union. *J. Wildl. Dis.* 43, S64–70.
- Reed, K.D., Meece, J.K., Henkel, J.S., Shukla, S.K., 2003. Birds, migration and emerging zoonoses: West Nile virus, lyme disease, influenza A and enteropathogens. *Clin. Med. Res.* 1, 5–12.
- Rushton, J., Viscarra, R., Guerne Bleich, E., McLeod, A., 2005. Impact of avian influenza outbreaks in the poultry sectors of five South East Asian countries (Cambodia, Indonesia, Lao PDR, Thailand, Viet Nam) outbreak costs, responses and potential long term control. *World's Poultry Sci. J.* 61, 491–514.
- Salman, M., Stark, K., Zepeda, C., 2003. Quality assurance applied to animal disease surveillance systems. *Rev. Sci. Tech. Off. Int. Epiz.* 22, 689–696.
- Thiermann, A.B., 2004. Emerging diseases and implications for global trade. *Rev. Sci. Tech. Off. Int. Epiz.* 23, 701–708.
- Thiermann, A.B., 2005. Globalization, international trade and animal health: the new roles of OIE. *Prev. Vet. Med.* 67, 101–108.
- Vallat, B., Pinto, J., Schdel, A., 2006. International organizations and their role in helping to protect the worldwide community against natural and intentional biological disasters. *Rev. Sci. Tech. Off. Int. Epiz.* 25, 163–172.
- Webster, R., Hulse, D., 2004. Microbial adaptation and change: avian influenza. *Rev. Sci. Tech. Off. Int. Epiz.* 23, 453–465.
- Zepeda, C., Salman, M., Thiermann, A., Kellar, J., Rojas, H., Willeberg, P., 2005. The role of veterinary epidemiology and veterinary services in complying with the World Trade Organization SPS agreement. *Prev. Vet. Med.* 67, 125–140.